Remote timed up and go evaluation from activities of daily living reveals changing mobility after surgery

Saporito, Salvatore; Brodie, Matthew Andrew; Delbaere, Kim; Hoogland, Jildou; Nijboer, Harmke; Rispens, Sietse Menno; Spina, Gabriele; Stevens, Martin; Annegarn, Janneke

Published in:
Physiological Measurement

DOI:
10.1088/1361-6579/ab0d3e

Remote Timed Up and Go evaluation from activities of daily living reveals changing mobility after surgery

Salvatore Saporito[1], Matthew Brodie[2], Kim Delbaere[2], Jildou Hoogland[3], Harmke Nijboer[4], Sietse Menno Rispens[1], Gabriele Spina[1], Martin Stevens[3], Janneke Annegarn[1]

1. Philips Research Europe, Eindhoven, the Netherlands, High Tech Campus 34 5656AE, Eindhoven, The Netherlands

2. Neuroscience Research Australia, University of New South Wales, Sydney, New South Wales, Australia, Wallace Wurth Building, UNSW Sydney NSW 2052 Australia

3. University of Groningen, University Medical Center Groningen, Department of Orthopedic Surgery, PO box 30.001 9700 RB Groningen, the Netherlands

4. Catharina Hospital Eindhoven, Eindhoven, the Netherlands, Michelangelolaan 2 5623 EJ Eindhoven

Correspondence to: salvatore.saporito@philips.com, High Tech Campus 34 5656AE, Eindhoven, The Netherlands. +31 (0) 6 18523859

Running title: TUG in daily life by wearable pendant

Author Contributions: SS and JA conceived the study. MB, KD, JH, HN, and MS substantially contributed to study design and to the acquisition of experimental data; SS, SR, and GS designed the computational framework and analysed the data. MB, KD, JH, MS, and HN contributed to the data interpretation. Manuscript was drafted by SS and JA, with critically important intellectual content from MB, KD, JH, HN, MS, and SR. All authors critically revised and approved the final manuscript.

Keywords: Activities of daily living; Hip Arthroplasty; Hospital discharge; Timed up and go; Wearable.

Abstract

Background. Mobility impairment is common in older adults and negatively influences the quality of life. Mobility level may change rapidly following surgery or hospitalization in the elderly.

The Timed Up and Go (TUG) is a simple, frequently used clinical test for functional mobility; however, TUG requires supervision from a trained clinician, resulting in infrequent assessments. Additionally, assessment by TUG in clinic settings may not be completely representative of the individual's mobility in their home environment.

Objective. In this paper, we introduce a method to estimate TUG from activities detected in free-living, enabling continuous remote mobility monitoring without expert supervision. The method is used to monitor changes in mobility following Total Hip Arthroplasty (THA).

Methods. Community-living elderly (n=239, 65-91 years) performed a standardized TUG in a laboratory and wore a wearable pendant device that recorded accelerometer and barometric sensor data for at least 3-days. Activities of daily living, including walks and sit-to-stand transitions, and their related mobility features were extracted and used to develop a regularized linear model for remote TUG test estimation. Changes in the remote TUG were evaluated in orthopaedic patients (n=15, 55-75 years), during 12-weeks following THA.

Results. In leave-one-out-cross-validation, a strong correlation ($\rho=0.70$) was observed between the new remote TUG and standardized TUG times. Test-retest reliability of 3-day estimates was high (ICC=0.94). Compared to 2-weeks post-THA, remote TUG was significantly improved at 6-weeks (11.7±3.9 s vs 8.0±1.8 s, $p<0.001$), with no further change at 12-weeks (8.1±3.9 s, $p=0.37$).

Conclusions. Remote TUG can be estimated in older adults using 3-days of ADLs data recorded using a wearable pendant. Remote TUG has discriminatory potential for identifying frail elderly and may provide a convenient way to monitor changes in mobility in unsupervised settings.

Introduction

Mobility is a key factor in healthy aging [1]. Mobility impairment is defined by a reduced postural balance and gait performance [2] that affects the ability to participate in activities of daily living (ADLs). Mobility impairment is common in the elderly population and it negatively influences the physical, psychological, and social wellness of older adults. Clinical review suggests primary care physicians should consider mobility as an essential aspect of the assessment of the frail older adult [3].

The Timed up and go test (TUG) is an established method to assess functional mobility, widely used in the literature as well as in the senior care industry [4, 5, 6]. The TUG test consists of getting up from a chair, walking 3-meters, turning 180 degrees, walking back to the chair, and sitting down [2]. The time to complete the task is recorded by a stopwatch and evaluated as test score; longer durations are associated with decreased mobility, and with adverse outcomes in the community-dwelling, as well as in institutionalized older adults [4].

The TUG-test has been widely cited in the literature as an outcome measure in a variety of conditions influencing mobility including stroke [7], chronic obstructive pulmonary disease [8] and Parkinson's disease [9]. Moreover, TUG has been recommended for assessment of walking and balance in guidelines for fall prevention [10]. The TUG test is generally carried out under supervision in clinic setting. Due to resources constraints, the TUG test is often administered at a one-time point only, which may limit its ability to detect changes in mobility over time.

Mobility may change rapidly in older adults, especially after hospital discharge [11] or during rehabilitation after orthopaedic surgery [12, 13, 14]. Whereas close supervision generally allows healthcare professionals to track changing mobility in hospitalized patients, tracking outpatients typically relies on less frequent physical examinations during clinical visits or self-reported questionnaires [14]. There is an unmet need for a cost-effective and objective way to remotely assess changes in mobility in unsupervised settings.

Wearable sensor-based measurements [15] may provide valid and reliable data about physical activity in different hospital discharged populations such as people who have had a stroke [16], joint replacement [14], or cardiac surgery [11]. Systematic reviews have concluded that objective physical activity data collected by body-worn sensors may be capable of predicting functional recovery post-operatively [17]. Indicators of functional mobility during daily life include cumulative physical activity such as step counts [14], the ability to perform different ADLs such as chair rise transfers [18] and walking quality [19].

However, it is currently unclear how best to summarize the variety of sensor-based measures into a single mobility outcome that clinicians are familiar with and that can be easily interpreted.

In this paper, a novel method for estimating mobility as TUG test value from ADLs recorded using a wearable pendant in free-living conditions is proposed. The method relies on the statistical relationships between mobility indicators measured in free-living conditions and the measured TUG test value, as observed in a population of older adults with a wide range of functional mobility. We aim also to gain insight into the clinical relevance of the proposed model; by examining free-living mobility data collected during 12 weeks in 15 patients recovering from Total Hip Arthroplasty (THA).

Methods

Study population

For the remote TUG model development, a population of older people (total n=319) was recruited from Sydney, Australia (n=159); Cologne, Germany (n=25); Valencia, Spain (n=21) and Eindhoven, The Netherlands (n=114) respectively (Table 1). Participants from Australia, Germany, and Spain took part in the SureStep, iStoppFalls [20] and StandingTall [21] randomized controlled trials to prevent falls. Participants from The Netherlands were discharged patients who had been admitted for non-surgical reasons (Eindhoven). Individuals were eligible if they were: (i) Aged 65 years or older; (ii) living independently; (iii) able to walk with or without a walking aid. Exclusion criteria were: (i) Major cognitive impairments; (ii) medical conditions preventing regular exercise.

For the assessment of changing mobility following surgery, volunteers participating in a post-THA rehabilitation program (n=20) were recruited at the Department of Orthopaedics, University Medical Centre Groningen, The Netherlands (Table 1). The THA participants underwent a THA as treatment for primary or secondary osteoarthritis. Patients were discharged a few days after surgery, in line with clinical practice in the Netherlands. Study participants followed a 12-week home-based exercise program with video instructions on a tablet PC. Patients performed strengthening and walking exercises at least five days a week [22] [23].

Daily activity assessments.

Participants were required to wear the Senior Mobility Monitor (SMM) (Philips Research, Eindhoven, The Netherlands) in their usual environment during their normal ADLs, with no restriction for indoor or outdoor activities. The SMM was worn on a lanyard in front of the chest. The SMM (Figure 1) is a pendant containing a tri-axial-accelerometer and a barometer (approximate size: 39 x 12 x 63 mm). Accelerations were recorded at a sampling frequency of 50 Hz and with a dynamic range of ±8 g; air pressure, negatively correlated with height, was recorded with a sampling frequency of 25 Hz.

Standardized TUG

All participants underwent a standardized TUG test at baseline before being supplied with an SMM. The TUG test consists of getting up from a chair, walking 3-meters, turning 180 degrees, walking back to the chair, and sitting down. Time taken was measured using a stopwatch. All participants were instructed to perform the TUG test at their normal pace.

Sensor data inclusion criteria

The minimal wearing time requirement for a day of sensor data to be considered valid was 8.0 hours. For the remote TUG model development, data inclusion criteria were having at least 3 valid days of activity data recorded within the 7 days following the standardized TUG. THA patients were included in the analysis if they had at least 30 valid days within the 12 weeks following hospital discharge.

Remote TUG evaluation model development

Remote mobility indicators were derived from the pendant sensor data. From walks detected in free-living using a dedicated pendant-based algorithm [24] several quantities were extracted: number of steps, walking duration, walking cadence, walking intensity [25], walking regularity (consistency of the stride-to-stride pattern) [26], walking symmetry (harmonic ratio) [27], walk instability, (Lyapunov exponent of stride-to-stride fluctuations [28] and sample entropy) [29], intensity characteristics of the dominant frequencies (peak frequency, amplitude, and width) [30]. Stride length was not estimated in this study.

The daily chair rise transition analysis included sit-to-stand transition duration, velocity, peak power, -maximum jerk, maximum acceleration and frequency [31] [32] [33]. Moreover, active and sedentary-bout lengths at multiple activity threshold levels were calculated [34]. Non-wearing periods were detected as the absence of movement for at least 15 minutes, based on sensor values, and excluded from the analysis.

Values measured from single events were calculated and aggregated using order statistics of the daily values (e.g. 90th-percentile), as well as measures of variability (e.g. interquartile range) [15]. The daily values were averaged over the week following the standardized TUG test and used to train the prediction model.

Remote TUG estimator training

Daily life remote mobility indicators, as well as TUG, were transformed to normal distributions using a box-cox transform. Remote mobility indicators with excellent test-retest reliability and significantly correlated with the standardized TUG were used for model training. A regularized linear model [35] was trained based on the mobility indicators that passed the correlation and test-retest reliability criteria. A regularized linear model was preferred over other model types for the interpretability of the final regression equation; coefficient shrinkage by means of L_1 and L_2 penalties were used to prevent over-fitting and to limit model complexity [35].

Nested cross-validation (CV) was used to evaluate the model performances (outer leave-one-out to determine the final model performance, inner 3-fold for model hyper-parameter optimization) [36]. Inner cross-validation was used to optimize hyper-parameters, using minimal hold-out error as model selection criterion.

The outer leave-one-out-cross-validation (LOOCV) was used to assess model performances which were evaluated in both (i) the overall population, and (ii) on a subsample with standardized TUG<20s. For LOOCV, at each iteration (i=239) of the model performance evaluation, the model was trained on data (n=238) from all but one participant (the hold-out participant). The model performance was then calculated using the combined predictions from all the held-out participants. Each model was trained on features selected using its respective training data only i.e. feature selection was performed independently and separately for each of the 239 models trained leaving one patient out, while the left-out patients were used to determine the reported model performances [37]. The final model performance was reported based on the aggregated hold-out results. Data used for the model performance evaluation were not used for model training.

The data from the post-THA participants were not used at any stage of the model training, constituting an independent validation set.

Remote TUG prediction of changing mobility following surgery

Data from the post-THA participants were not used for model development but kept separate to evaluate the remote TUG predictions on unseen data. A 3-days sliding-window was used to evaluate remote TUG. An example of mobility indicators used for model training is shown in Figure 2.

Statistical analyses

Correlations between standardized TUG and individual mobility indicators were assessed by Spearman correlation coefficient; a significance level of 0.05 was used, with Benjamini-Hochberg correction for control of false discovery rate [38].

Test-retest reliability of each remote mobility indicator was calculated using two estimates of the same indicator, each obtained averaging two subsets of 3 daily values (as shown in Figure 2). Only subjects with sufficient (6 or more) valid days were used for this purpose. The degree of agreement between the two independent 3-days averages was quantified as intra-class correlation coefficient (ICC) for absolute agreement of indicators, which are average of k independent tests (ICC(2,k)), and by the corresponding confidence intervals for the ICC [39]. ICC greater than 0.9 was used as a common measurement threshold for individual clinical decision-making [40].

The agreement between the remote TUG and standardized TUG was measured by Spearman's correlation coefficient. The performance of the model was measured as the median absolute error between remote TUG and standardized TUG. Moreover, the ability of the model to differentiate between participants with a TUG of above 10 s and below 10 s was quantified as area under the receiver-operator characteristic curve (AUC-ROC).

The performances of the remote TUG prediction model were evaluated using LOOCV in the complete training dataset (n=239), and in participants with a standardized TUG<20.0 only (n=231). The rationale for presenting results with and without the n=8 subjects with TUG>20 s is that these subjects are effectively outliers with respect to the TUG distribution in the training population, and we hypothesised they might heavily influence global performance measures. Therefore, a sensitivity analysis was conducted to determine the effect of these outliers on the model performance and both values are reported in the results. For the THA population, values of remote TUG in the different weeks were compared using analysis of variance (ANOVA) with post-hoc paired t-tests and Bonferroni correction for multiple comparisons. A corrected p-value smaller than 0.05 was considered significant.

Results

In total, 339 participants were recruited (of which 20 THA participants). 254 participants passed the data analysis inclusion criteria (of which 15 were part of the THA study). Participants were excluded if they did not wear the SMM in the week following the standardized TUG test (n=7); others (n=78) were excluded because they did not meet the minimum wearing time data validity criteria (3 valid days for the model development population, 30 days for post-THA population). For the remote TUG model development and cross-validation, 239 participants wore the SMM for 8 hours or more for at least 3-days resulting in 1302 days of data. Demographics for the included participants are presented in Table 1.

An overview of the features used for TUG model, together with their correlation with TUG and reliability values are shown in Table 2. The performance of the remote TUG prediction model was evaluated using LOOCV in the complete training dataset (n=239), and in participants with a standardized TUG<20.0 only (n=231, results between brackets). The results for the latter sensitivity analysis are shown between brackets. The Spearman correlation coefficient between the remote TUG and the standardized TUG was 0.70 (0.67), while the median absolute error was 1.4 s (1.3 s) and the mean absolute error was 2.1 s (1.7 s). The AUC-ROC curve for the classification of the participants with standardized TUG>10s was 0.89. A comparison between standardized and remote TUG by Bland-Altman [41] analysis is shown (Figure 3).
The final remote TUG prediction model was based on six mobility indicators including the categories of walking quantity, walking quality, chair rise quality, active and inactive bout durations. The remote TUG demonstrated high test-retest in the model development population (ICC=0.94) between repeated assessments based on different 3-day windows of data (Figure 2B).

The remote TUG outcomes for the 15 included subjects during rehabilitation after THA are shown in Figure 4. In the first 6 weeks, the remote TUG reduced significantly indicating improved mobility as participants recovered from surgery ($p<0.001$). Remote TUG was significantly different in the THA population when comparing week 2 (11.7 ± 3.9 s) after surgery with week 6 (8.0 ± 1.8 s) and with week 12 (8.1 ± 3.9 s) after surgery (both $p<0.001$), while the differences were not significant when comparing week 6 and week 12 ($p=0.37$), as shown in Figure 5. The Spearman correlation coefficient between baseline pre-operative TUG and remote TUG in THA was $\rho=-0.04$ (n=15, $p=0.87$), $\rho=-0.11$ (n=15, $p=0.68$), and $\rho = 0.22$ (n=12, $p=0.48$) at weeks 2, 6, and 12, respectively.

Discussion

In this work, a model was developed to estimate TUG remotely using 3-days of wearable pendant sensor recordings during unsupervised ADLs. We found strong correlation between the standardized TUG and the remote TUG estimated from free living data. In addition, we showed that the remote TUG estimator was sensitive to improvements in mobility after hospital discharge following THA. As TUG is a widely accepted method to assess functional mobility [4], relating the aggregated sensor-based mobility indicators to standardized TUG times may aid clinical interpretability.

Remote TUG performance and reliability

The aim of this study was to introduce a novel method for remote assessment of the standardized TUG test time from unsupervised ADLs, in order to provide a clinically relevant measure of functional mobility in free-living conditions. In this study, a strong correlation between the standardized TUG and the remotely predicted TUG was found (0.70), indicating that there was agreement between sensor-based estimation and standardized test value. The remotely predicted TUG demonstrated the good discriminatory power (AUC 0.89) to identify subjects with TUG>10 seconds, which has been suggested by literature as the cutoff value for frailty [42] [43].

There is ongoing debate in the scientific community over cut-off values for standardized TUG and alternative cut-off values have been proposed [44, 45]; conclusions regarding which value should be used as cut-off in clinical practice are outside the scope of this manuscript.

Reliability and data validity criterion

The high test-retest reliability observed for the remote TUG confirms (ICC=0.94) that a monitoring period of 3 days, each day with sufficient wearing time, is sufficient for assessment of daily-life mobility in unsupervised settings. This suggests the remote TUG may be reliable enough to be used for clinical decisions at the individual level. The 3-days duration is in agreement with previous studies on remote monitoring of gait quality, gait intensity and gait quantity [46, 47] in community-living older people. Regarding the minimum amount of wearing time to consider a day to be valid for the analysis (8 hours), this value corresponds to roughly 50% of a typical awake time. Although we did not evaluate the sensitivity of the results to this parameter in this study, similar values have been used in the literature [48] [49].In this study, this amount of time was shown to be sufficient to record events which lead to reliable daily statistics estimates for quality of motion indicators. Further research would be needed to assess the data validity criteria for other use cases. Our study confirmed that pendant-based mobility monitoring for three days is sufficient for a reliable remote TUG estimate (ICC=0.94) in community-dwelling older populations, including subjects recently discharged from the hospital.

Standardized TUG versus remote TUG.

Differences were observed between the standardised TUG and remote TUG times.

We attribute part of the error magnitude to the relatively large errors in subjects with standardised TUG greater than 15 s, which were not adequately represented in the used training set (Figures 2 and 3A). This influence on the final metrics of large outliers is reflected by a small median absolute error of 1.4 s (1.3 s) compared to a mean absolute error of 2.1 s (1.7 s).

Further differences are due to the fact that while_standardized TUG assessments reflect function mobility under "*optimum conditions"*, the remotely predicted TUG may reflect "*actual"* functional mobility (defined as walking and balance performance in "*free-living"* conditions). Fundamental differences between these two assessments include:

1. *Assessments in free-living conditions versus assessment in a standardized environment*

Performances in standardized settings may not replicate ecological conditions [5]. Standardized test conditions may result in nearly maximal performance with limited variability between repetitions [50]. Remote mobility assessments may include evaluation of performances at different effort levels [18]. The clinical relevance of the spread of performances during ADLs has previously been demonstrated by associations with falls risk, psychological, cognitive, and sensorimotor health factors [51, 52].

2. *Quality of movement versus quantity of movement*

Similar to existing instrumented TUG tests [53, 5], the new remotely predicted TUG has the advantage that, in addition to movement intensity, the quality of movements may also be assessed. However, our approach is different from others in the literature [54]**,** [55], which used motion sensors as improved (respect to standard stopwatch) measurement system for a supervised/unsupervised/self-administered standardized test. The remotely predicted TUG times were based on multiple mobility indicators including gait intensity, gait quantity, and gait quality, whereas for the standardized TUG, only total time taken to complete the task is generally reported.

3. *Continuous unsupervised assessment versus single time point supervised assessment*

Awareness of being in a “test” situation, often results in better performances, known as or the “white coat effect” [5] [55]. The remotely predicted TUG times were based on unsupervised activities recorded over 3 days, therefore they may be less dependent on short-term influences such as footwear [56] or medication intake, and assessor influences such as the choice of the chair or instructions provided. From an access to healthcare perspective, unsupervised assessments may be less affected by the availability of clinicians, proximity to a clinic and cost constraints.

Remote TUG evaluation post-THA surgery

Widespread rapid recovery protocols after total hip arthroplasty (THA) have reduced the length of hospital stay to a few days [57]. As patients leave the hospital while their mobility is still impaired, it would be beneficial and valuable for the formal and informal caregiver to remotely monitor post-operative progress in outpatient settings.

The few studies that have reported objectively assessed post-operative mobility in free-living have generally focused on the aggregated amount of daily physical activity performed (e.g. total steps/day [13, 14, 12]). Comparatively, in this study, remote TUG estimates were based on a quantitative characterization of different ADLs (e.g. peak acceleration during a sit to stand, walk intensity, distribution of sedentary time, ….) derived from the raw sensor data.

Consistent across all patients (Figure 5) was a rapid improvement in mobility (reduced remote TUG times) between weeks two and six (from 11.7 ± 3.9 s to 8.0 ± 1.8 s) with little change observed between week six and twelve. Monitoring changing mobility in unsupervised settings could help provide tailored therapeutic guidance aimed at optimizing adherence to interventions [5]. With respect to post-THA recovery, remote TUG times could be used by the caregiver to tailor the duration or intensity of the rehabilitation, or by the patient to support self-management and goal setting [58]. The rapid improvements in mobility between weeks two and six indicate the importance of using this window of opportunity for targeted rehabilitation following hospital discharge.

Limitations and future research

The model was developed using sensor-based measurements alone, i.e. without including any subject anthropometric, clinical or demographic characteristics. The use of additional external information e.g. subject height or gender, could have improved the presented model estimates [59]. However, in this study, only sensor-based features were used to demonstrate the feasibility of solely sensor-based remote TUG estimation.

We acknowledge certain limitations. Although precautions (including LOOCV) were taken to ensure the remote TUG model was not over-fitted, generalizability may be limited by the demographics of the people included in the training dataset. Specifically, participants were generally independent-living older people of European ethnicity. The THA patient population of the validation set was not directly comparable with the training set populations. When considering the clinical application of the method in the real world, data from specific target populations should be included in the training set. However, in this first feasibility study context, we have chosen to use the THA as a validation set only. When moving to real-world clinical application, more data from THA and other clinical populations would be included in the expanded training set for developing a more generalizable final production model. Specifically, future remote TUG algorithm development should include the additional recruitment of subjects with standardised TUG times greater than 20 s. Future research would be required to assess the cost-effectiveness of remote TUG as (i) an objective measure to telemonitor progress in response to a certain treatment or intervention in outpatient settings and (ii) to close the loop by allowing caregivers to tailor the type, duration, or intensity of an intervention based on continuous assessments of changing mobility and compliance with prescribed exercises (e.g. post-operative rehabilitation).

Several studies have proposed and shown the additional value of instrumenting laboratory physical tests with wearable sensors [60]**,** [54] [61] [62] [63] [64]. It is a limitation of this study that the participants did not wear the sensor during baseline TUG assessments. Training a TUG model on this more specific data could have yielded improved algorithm accuracy, i.e. closer numerical value to the actual standardized TUG value. This would have provided a greater understanding of the upper limit for expected accuracy (under ideal conditions) when a TUG model would be developed on the more variable daily-life data. In future studies, it would also be interesting to compare remote TUG and instrumented TUG, for instance, to evaluate the relationship between sub-phases of the standardized TUG [37] with specific mobility features.

In conclusion, a new method was developed to evaluate TUG remotely from unsupervised ADLs recorded using a wearable pendant. The method was used to continuously monitor changes in mobility following THA. The remotely predicted TUG was strongly correlated with standardized TUG times and had high test re-test reliability, which may make it suitable for clinical decisions at the individual level.

Due to the universal use of TUG in the healthcare and senior care industry, the new remotely predicted TUG may be easily interpretable for healthcare professionals. Future studies should focus on how ongoing remote TUG prediction might improve clinical decision-making and assist in developing more personalized interventions for older adults.

Acknowledgments and funding.

SS, SR, GS, JA are employed by Philips Research. Philips Research was the listed as industry partner of KD within NHMRC career development fellowship. The remaining authors declare that they have no further competing interests. The iStoppFalls project has received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement [287361]. The Australian arm was funded by an Australian National Health and Medical Research Council (NHMRC) EU collaboration grant [#1038210]. The StandingTall project was supported by NHMRC [APP1084739]. The study at University Medical Center Groningen was supported by "Stichting Anna Fonds NOREF" and "ZonMW" [40-00812-98-09014].

The content of the manuscript does not represent the opinion of the European Community or NHMRC. The funding sources had no influence on any aspect of this study.

References

[1] W. A. Satariano, J. M. Guralnik, R. J. Jackson, R. A. Marottoli, E. A. Phelan and T. R. Prohaska, "Mobility and aging: new directions for public health action," *American Journal of Public Health,* vol. 102, no. 8, pp. 1508-1515, 2012.

[2] D. a. S. R. Podsiadlo, "The timed "Up & Go": a test of basic functional mobility for frail elderly persons," *Journal of the American Geriatrics Society,* vol. 39, no. 2, pp. 142-148, 1991.

[3] C. J. Brown and K. L. Flood, "Mobility Limitation in the Older Patient. A clinical review," *Journal of the American Medical Association,* vol. 310, no. 11, pp. 1168-77, 2013.

[4] H. A. Bischoff, H. B. Stähelin, A. U. Monsch, M. D. Iversen, A. Weyh, M. von Dechend, R. Akos, M. Conzelmann, W. Dick and R. Theiler, "Identifying a cut-off point for normal mobility: a comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women," *Age and Ageing,* vol. 3, no. 1, p. 315–320, 2003.

[5] G. Sprint, D. J. Cook and D. L. Weeks, "Towards Automating Clinical Assessments: A Survey of the Timed Up and Go (TUG)," *IEEE Rev Biomed Eng,* vol. 8, pp. 64-77, 2015.

[6] R. W. Bohannon, "Reference values for the timed up and go test: a descriptive meta-analysis.," *Journal of geriatric physical therapy,* vol. 29, no. 2, pp. 64-68, 2006.

[7] S. S. Ng and C. W. Hui-Chan, "The timed up & go test: its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke," *Archives of physical medicine and rehabilitation ,* vol. 86, no. 8, pp. 1641-1647, 2005.

[8] M. A. Al Haddad, J. Michelle, H. Samia and E. B. Charlotte, "Role of the Timed Up and Go Test in patients with chronic obstructive pulmonary disease," *Journal of cardiopulmonary rehabilitation and prevention,* vol. 36, no. 1, pp. 49-55, 2016.

[9] E. L. Stegemöller, J. Nocera, I. Malaty, M. Shelley, M. S. Okun and C. J. Hass, "Timed up and go, cognitive, and quality-of-life correlates in Parkinson's disease," *Archives of Physical Medicine and Rehabilitation,* vol. 95, no. 4, pp. 649-655, 2014.

[10] R. A. M. Kenny, L. Z. Rubenstein, M. E. B. K. Tinetti, K. A. Cameron, E. A. Capezuti and D. P. John, "Summary of the updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons," *Journal of the American Geriatrics Society,* vol. 59, no. 1, pp. 148-157, 2011.

[11] D. J. Cook, J. E. Thompson, S. K. Prinsen, J. A. Dearani and C. Deschamps, "Functional Recovery in the Elderly After Major Surgery: Assessment of Mobility Recovery Using Wireless Technology,"

Annals of thoracic surgery, vol. 96, no. 3, p. 1057–1061, 2013.

[12] N. Mehta, C. Steiner, K. Fields, D. Nawabi and S. Lyman, “Using mobile tracking technology to visualize the trajectory of recovery after hip arthroscopy: a case report.,” *HSS Journal,* vol. 13, no. 2, pp. 194-200, 2017.

[13] I. de Groot, H. J. Bussmann, H. J. Stam and J. A. Verhaar, “Small increase of actual physical activity 6 months after total hip or knee arthroplasty,” *Clin Orthop Relat Res. ,* vol. 466, no. 9, p. 2201–2208, 2008.

[14] P. Toogood, M. P. Abdel, J. A. Spear, S. M. Cook, D. J. Cook and M. J. Taunton, “The monitoring of activity at home after total hip arthroplasty,” *Bone Joint Journal,* vol. 98, no. 11, pp. 1450-1454, 2016.

[15] T. R. S. J. N. M. L. N. Shany, “Sensors-based wearable systems for monitoring of human movement and falls,” *IEEE Sensors Journal,* vol. 12, no. 3, pp. 658-670, 2012.

[16] N. Gebruers and e. al, “Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures. Archives of physical medicine and rehabilitation,” *Archives of physical medicine and rehabilitation,* vol. 91, no. 2, pp. 288-297, 2010.

[17] A. Abeles, R. M. Kwasnicki, C. Pettengell, J. Murphy and A. Darzi, “The relationship between physical activity and post-operative length of hospital stay: A systematic review,” *International Journal of Surgery,* vol. 44, p. 295–302, 2017.

[18] W. Zhang, R. G. H. Regterschot, H. Geraedts, H. Baldus and W. Zijlstra, “Chair Rise Peak Power in Daily Life Measured with a Pendant Sensor Associates with Mobility, Limitation in Activities and Frailty in Old People.,” *IEEE J Biomed Health Inform,* vol. 21, no. 1, pp. 211-2017, 2015.

[19] M. A. Brodie, Y. Okubo, J. Annegar, R. Wieching, S. R. Lord and K. Delbaere, “Disentangling the health benefits of walking from increased exposure to falls in older people using remote gait monitoring and multi-dimensional analysis.,” *Physiological measurement,* vol. 38, no. 1, pp. 45-62, 2017.

[20] Y. J. Gschwind, S. Eichberg, H. R. Marston, A. Ejupi, H. de Rosario, M. Kroll, M. Drobics, J. Annegarn, R. Wieching, S. R. Lord, K. Aal and K. Delbaere, “ICT-based system to predict and prevent falls (iStoppFalls): study protocol for an international multicenter randomized controlled trial,” *BMC Geriatrics,* vol. 14, no. 91, 2014.

[21] K. Delbaere and e. al., “Evaluating the effectiveness of a home-based exercise programme delivered through a tablet computer for preventing falls in older community-dwelling people over 2 years: study protocol for the StandingTall randomised controlled trial,” *BMJOpen,* vol. 5, p. e009173, 2015.

[22] H. Jildou, W. Annet, M. Tjerk, G. C. LE, D. Baukje, W. P. Zijlstra, A. Janneke, I. Francisco, Z. Wiebren and S. Martin, “Feasibility and Patient Experience of a Home-Based Rehabilitation Program Driven by a Tablet App and Mobility Monitoring for Patients After a Total Hip

Arthroplasty," *JMIR,* vol. 10.2196/10342, p. In press, 2018.

[23] Wijnen et al., "Effectiveness of a home-based rehabilitation program driven by a tablet application and distant coaching: a comparison with usual care in the Netherlands for patients following total hip arthroplasty," *Under Review.*

[24] J. Annegarn, W. R. T. Ten Kate and N. G. P. Den Teuling, "Improvements in the detection of walking in measurements of the movement of a user US20170000384A1". US Patent 15/114,095, 5 January 2017.

[25] C. V. Bouten, K. R. Westerterp, M. Verduin and J. D. Janssen, "Assessment of energy expenditure for physical activity using a triaxial accelerometer," *Medicine and science in sports and exercise,* vol. 12, no. 26, pp. 1516-1523, 1994.

[26] R. Moe-Nilssen and J. L. Helbostad, "Interstride trunk acceleration variability but not step width variability can differentiate between fit and frail older adults," *Gait & posture,* vol. 21, no. 2, pp. 164-170, 2005.

[27] J. L. Bellanca, K. A. Lowry, J. M. VanSwearingen, J. S. Brach and M. S. Redfern, "Harmonic ratios: a quantification of step to step symmetry," *Journal of biomechanics,* vol. 46, no. 4, pp. 828-831, 2013.

[28] M. Toebes, M. Hoozemans, R. Furrer, J. Dekker and J. van Dieën, "Local dynamic stability and variability of gait are associated with fall history in elderly subjects," *Gait & posture,* vol. 36, no. 3, pp. 527-531, 2012.

[29] F. Riva, M. Bisi and R. Stagni, "Gait variability and stability measures: Minimum number of strides and within-session reliability," *Computers in biology and medicine,* vol. 50, pp. 9-13, 2014.

[30] B. Caby, S. Kieffer, M. Saint Hubert, G. Cremer and B. Macq, "Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry," *Biomedical engineering online,* vol. 10, no. 1, p. 1, 2011.

[31] M. C. Boonstra, P. J. Schwering, M. C. De Waal-Malefijt and N. Verdonschot, "Sit-to-stand movement as a performance-based measure for patients with total knee arthroplasty.," *Physical therapy,* vol. 90, no. 2, pp. 149-156, 2010.

[32] U. Lindemann, H. Claus, M. Stuber, P. Augat, R. Muche, T. Nikolaus and C. Becker, "Measuring power during the sit-to-stand transfer," *Eur. J. Appl. Physiol.,* vol. 89, no. 5, p. 466–470, 2003.

[33] R. G. H. Regterschot, M. Folkersma, W. Zhang, H. Baldus, M. Stevens and W. Zijlstra, "Sensitivity of sensor-based sit-to-stand peak power to the effects of training leg strength, leg power and balance in older adults," *Gait & posture,* vol. 39, no. 1, pp. 303-307, 2014.

[34] S. Bae, A. K. Dey and C. Low, "Using passively collected sedentary behavior to predict hospital readmission," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and*

Ubiquitous Computing, 2016.

[35] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 67, no. 2, pp. 301-320, 2005.

[36] G. C. Cawley and N. L. C. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *Journal of Machine Learning Research,* vol. 11, pp. 2079-2107, 2010.

[37] P. Smialowski, D. Frishman and S. Kramer, “Pitfalls of supervised feature selection,” *Bioinformatics,* vol. 26, no. 3, p. 440–443, 2010.

[38] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the royal statistical society,* vol. 57(1), pp. 289-300, 1995.

[39] K. O. McGraw and S. P. Wong, “Forming inferences about some intraclass correlation coefficients,” *Psychological Methods,* vol. 1(1), pp. 30-46, 1996.

[40] J. Kottner, L. Audigé, S. Brorson, A. Donner, B. Gajewski, A. Hróbjartsson, C. Roberts, M. Shoukri and D. Streiner, “Guidelines for reporting reliability and agreement studies (GRRAS) were proposed,” *International journal of nursing studies,* vol. 48, no. 6, pp. 661-671, 2011.

[41] J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” *The Lancet,* vol. 8476, pp. 307-310, 1986.

[42] G. Savva, O. Donoghue, F. Horgan, C. O'Regan, H. Cronin and R. Kenny, “Using timed up-and-go to identify frail members of the older population.,” *J Gerontol A Biol Sci Med Sci,* vol. 68, no. 4, pp. 441-6, 2013 .

[43] A. Clegg, L. Rogers and J. Young, “Diagnostic test accuracy of simple instruments for identifying frailty in community-dwelling older people: a systematic review.,” *Age Ageing,* vol. 44, no. 1, pp. 148-52, 2015 .

[44] E. Barry, R. Galvin, C. Keogh, F. Horgan and T. Fahey, “Is the Timed up and Go Test a Useful Predictor of Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-Analysis,” *BMC Geriatr.,* vol. 14, no. 1, 2014.

[45] G. Thrane, R. Joakimsen and E. Thornquist, “The Association between Timed up and Go Test and History of Falls: The Tromso Study,” *BMC Geriatrics,* vol. 7, no. 1, 2007.

[46] A. Weiss, M. Brozgol, M. Dorfman, T. Herman, S. Shema, N. Giladi and J. Hausdorff, “Does the evaluation of gait quality during daily life provide insight into fall risk? novel approach using 3 d accelerometer recordings,” *Neurorehabil. Neural Repair ,* vol. 27 , p. 742–52, 2013 .

[47] M. A. Brodie, S. R. Lord, M. J. Coppens, J. Annegarn and K. Delbaere, “Eight weeks remote monitoring using a freely worn device reveals unstable gait patterns in older fallers,” *IEEE Trans*

Biomed Eng, vol. 62, no. 11, pp. 2588-94, 2015.

[48] J. A. Schrack, R. Cooper, A. Koster, E. J. Shiroma, J. M. Murabito, W. J. Rejeski, L. Ferrucci and T. B. Harris, “Assessing daily physical activity in older adults: unraveling the complexity of monitors, measures, and methods,” *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* vol. 71, no. 8, pp. 1039-1048, 2016.

[49] M. J. Chinapaw, M. de Niet, M. Verloigne, I. B. J. De Bourdeaudhuij and T. M. Altenburg, “From sedentary time to sedentary patterns: accelerometer data reduction decisions in youth,” *PLoS One,* vol. 9, no. 11, p. e111205., 2014.

[50] M. Brodie, M. Coppens, S. Lord, N. Lovell, Y. Gschwind, S. Redmond, M. Del Rosario, K. Wang, D. Sturnieks, M. Persiani and K. Delbaere, “Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different,” *Medical & biological engineering & computing.,* vol. 54, no. 4, pp. 663-74, 2016.

[51] J. M. Hausdorff, “Gait variability: methods, modeling and meaning,” *J Neuroengineering Rehabil,* vol. 2, p. 19, 2005.

[52] M. A. Brodie, M. J. Coppens, A. Ejupi, Y. J. Gschwind, J. Annegarn, D. Schoene, R. Wiesching and S. R. D. K. Lord, “Comparison between clinical gait and daily-life gait assessments of fall risk in older people,” *Geriatrics and Gerontology International,* 2017.

[53] A. Salarian, F. Horak, C. Zampieri, P. Carlson-Kuhta, J. Nutt and K. Aminian, “iTUG, a sensitive and reliable measure of mobility.,” *IEEE Trans Neural Syst Rehabil Eng,* vol. 18, no. 3, pp. 303-10, 2010.

[54] L. Palmerini, S. Mellone, L. Rocchi and L. Chiari, “Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test,” in *Engineering in Medicine and Biology Society*, 2011.

[55] T. Frenken, B. Vester, M. Brell and A. Hein, “aTUG: Fully-automated timed up and go assessment using ambient sensor technologies,” in *InPervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference*, 2011.

[56] S. A. Arnadottir and V. S. Mercer, “Effects of footwear on measurements of balance and gait in women between the ages of 65 and 93 years,” *Physical Therapy,* vol. 80, no. 1, pp. 17-27, 2000.

[57] Y. M. den Hartog, N. M. C. Mathijssen and S. B. Vehmeijer, “Reduced length of hospital stay after the introduction of a rapid recovery protocol for primary THA procedures: A retrospective cohort study with 1,180 unselected patients,” *Acta orthopaedica,* vol. 84, no. 5, pp. 444-447, 2013.

[58] M. S. Austin, B. T. Urbani, A. N. Fleischman, N. D. Fernando, J. J. Purtill, W. J. Hozack, J. Parvizi and R. H. Rothman, “Formal Physical Therapy After Total Hip Arthroplasty Is Not Required: A Randomized Controlled Trial,” *JBJS,* vol. 99, no. 8, pp. 648-655, 2017.

[59] B. R. Greene, S. J. Redmond, Caulfield and B., "Fall Risk Assessment through Automatic Combination of Clinical Fall Risk Factors and Body-Worn Sensor Data," *IEEE Journal of Biomedical and Health Informatics,,* vol. 1, no. 1, p. 21, 2016.

[60] T. Frenken, B. Vester, M. Brell and A. Hein, "aTUG: Fully-automated timed up and go assessment using ambient sensor technologies.," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 5th International Conference on*, 2011.

[61] A. Galan-Mercant and A. I. Cuesta-Vargas, "Clinical frailty syndrome assessment using inertial sensors embedded in smartphones.," *Physiological measurement,* vol. 36, no. 9, p. 1929, 2015.

[62] W. A, T. Herman, M. Plotnik, M. Brozgol, N. Giladi and J. M. Hausdorff, "An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers," *Pysiological Measurement,* vol. 32, no. 12, p. 2003, 2011 .

[63] B. R. Greene, E. P. Doheny, R. A. Kenny and B. Caulfield, "Greene, B. R., Doheny, E. P., Kenny, R. A., & Caulfield, B. (2014). Classification of frailty and falls history using a combination of sensor-based mobility assessments. Physiological measurement, 35(10), 2053.," *Physiological measurement,* vol. 35, no. 10, p. 2053, 2014.

[64] A. Godfrey, J. Lara, C. Munro, C. Wiuff, S. Chowdhury, S. Del Din, A. Hickey, J. Mathers and L. Rochester, "Instrumented assessment of test battery for physical capability using an accelerometer: a feasibility study," *Physiological measurement,* vol. 36, no. 5, p. N71, 2015.

Tables

	Model development	Model evaluation
N	319	20
with sufficient wearing time	239	15
Gender	144:F; 95:M	10: F; 5: M
Age [years]	75.2±6.1* [65- 91]	63.8±6.5** [55-75]
Height [cm]	165.7±9.7* [143- 193]	174.6 ±7.2** [162- 186]
Weight [Kg]	74.3±16.1* [39 - 128]	74.8±10.4** [55-96]
BMI [Kg/m^2]	27.0±5.1* [15.6 - 52.6]	24.5±3.0** [19.6 - 31.6]
Standardized TUG [s]	9.79±4.41* [4.03 – 35.45]	10.46±1.83*† [7.58 - 15.53]
Five times Sit to stand Test [s]	14.38 ± 5.46 [5.15 – 35.49] (not available in n=8)	-
10 meters walk test [s]	8.78±3.30* [5.13 - 36.50] (not available in n=35)	-

Table 1: Demographic information and mobility values for included study participants. Values are expressed as mean values ± the corresponding standard deviation for normally distributed parameters, or median ± interquartile range for non-normally-distributed parameters [min – max values in the population]. *TUG= timed up and go. *= Normally distributed (p>0.05 for Shapiro-Wilk test) **= Non-normally distributed (p<0.05 for Shapiro-Wilk test).* † = pre-operative values

	n	Value	Correlation with TUG		Reliability	
Duration of active periods						
Sum active time	239	1480.49±608.12* [239.92 - 3103.98]	-0.5 [-0.6 , -0.39]	3.36E-16	n=134	0.91 [0.87 , 0.94]
Sum sedentary time	239	3806.04±1129.66* [1889.76 - 9742.39]	0.29 [0.16 , 0.41]	8.28E-06	n=134	0.65 [0.51 , 0.76]
Features derived from chair rise events						
Peak Jerk	239	0.54±0.13* [0.21 - 1.23]	-0.17 [-0.3 , -0.04]	8.54E-03	n=133	0.88 [0.84 , 0.92]
Peak Power	239	4.85±2.38** [1.28 - 9.69]	-0.07 [-0.2 , 0.08]	3.48E-01	n=133	0.94 [0.91 , 0.96]
Peak vertical acceleration	239	1.50±0.33* [0.78 - 2.47]	-0.16 [-0.29 , -0.03]	1.38E-02	n=133	0.94 [0.91 , 0.96]
Ratio acceleration time to total time	239	1.44±0.44** [0.76 - 2.21]	-0.16 [-0.29 , -0.03]	1.68E-02	n=133	0.95 [0.92 , 0.96]
Total Jerk	239	3.90±1.13* [1.47 - 9.44]	-0.22 [-0.35 , -0.09]	6.78E-04	n=133	0.91 [0.87 , 0.94]
Features derived from walk events						
Average stride time	239	1.17±0.09* [0.93 - 1.61]	0.3 [0.18 , 0.42]	2.39E-06	n=134	0.96 [0.95 , 0.98]
Fractal dimension	239	1.46±0.17* [1.06 - 2.06]	0.19 [0.06 , 0.32]	3.78E-03	n=134	0.95 [0.92 , 0.96]
Harmonic Ratio	239	5.13±2.50* [0.95 - 16.22]	-0.41 [-0.51 , -0.29]	9.38E-11	n=134	0.93 [0.89 , 0.95]
High frequency percentage power	239	8.44±4.44* [1.99 - 35.26]	-0.03 [-0.17 , 0.11]	6.58E-01	n=134	0.93 [0.91 , 0.95]
Intensity	239	4.35±2.10* [0.34 - 9.91]	-0.65 [-0.72 , -0.56]	2.25E-29	n=134	0.95 [0.93 , 0.97]
Low frequency percentage power	239	3.26±1.86* [1.00 - 15.05]	0.39 [0.27 , 0.5]	7.73E-10	n=134	0.93 [0.9 , 0.96]
Lyapunov exponents	239	0.24±0.10** [0.03 - 0.49]	-0.21 [-0.33 , -0.07]	1.65E-03	n=134	0.94 [0.91 , 0.96]
FT peak location	239	1.83±0.29* [1.32 - 3.62]	-0.13 [-0.26 , 0.01]	4.68E-02	n=134	0.9 [0.86 , 0.93]
FT peak amplitude	239	409.90±158.81* [181.24 - 1047.26]	-0.64 [-0.71 , -0.55]	4.42E-28	n=134	0.93 [0.9 , 0.95]
FT peak width	239	20.49±1.03* [19.00 - 23.10]	0.28 [0.15 , 0.4]	1.76E-05	n=134	0.78 [0.69 , 0.85]
Sample entropy	239	0.68±0.17** [0.43 - 0.99]	0.28 [0.15 , 0.4]	1.83E-05	n=134	0.96 [0.94 , 0.97]
Stride regularity	239	0.65±0.12* [0.34 - 0.93]	-0.37 [-0.48 , -0.25]	5.92E-09	n=134	0.92 [0.88 , 0.94]

Table 2: Mobility features measured using senior mobility monitor in the population used for model development. For each feature, the mean value obtained averaging all events in one day is presented. Values are expressed as mean values ± the corresponding standard deviation for normally distributed parameters, or median ± interquartile range for non-normally-distributed parameters [min – max values in the population]. Correlation coefficient is expressed as Spearman's correlation coefficient, [lower – upper bound for the 95% confidence intervals] and corresponding p-value. Reliability is quantified intra-class correlation coefficient (ICC) for absolute agreement of indicators, which are average of k independent tests ICC (2,k) [lower – upper bound for the 95% confidence intervals] **= Normally distributed (p>0.05 for Shapiro-Wilk test) **= Non-normally distributed (p<0.05 for Shapiro-Wilk test). TUG= timed up and go. FT= Fourier transform*

Figure legends

Figure 1. Senior Mobility Monitor (SMM). The device was worn in front of the chest using a provided necklace. The SMM contains a tri-axial-accelerometer and a barometer (approximate device size: 39 x 12 x 63 mm).

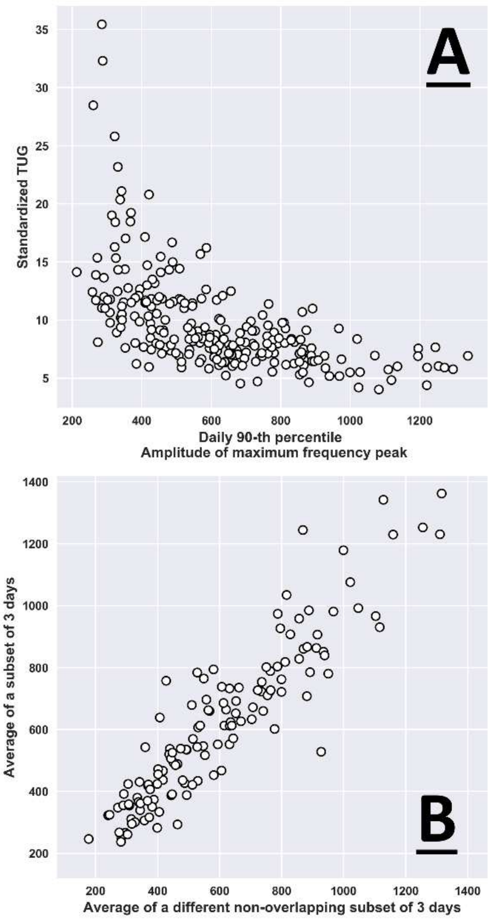

Figure 2. Scatter plots showing performance of a remote mobility indicator (daily inter-quartile range of the walking spectral peak amplitude). A: remote mobility assessment type against standardized TUG time and B: average values remote mobility assessment type for two distinct 3 days subsets within the baseline week for each subject. Reliability was quantified comparing the two 3-days averages. Spearman correlation coefficient ρ=-0.67 (n=239), ICC(2,k)=0.93 (n=134, subset with at least 6 valid days).

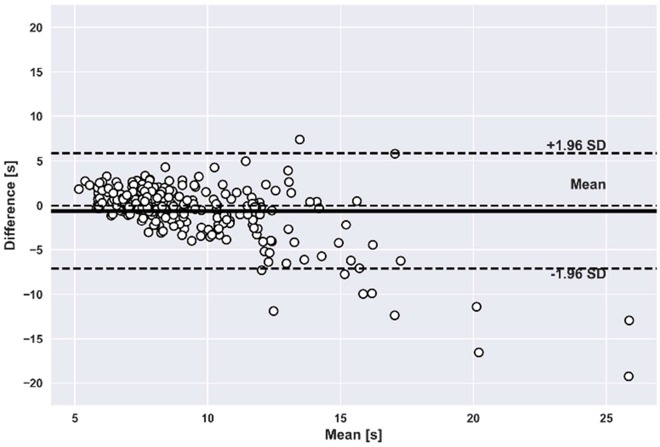

Figure 3. Bland-Altman plot comparing Remote TUG and standardized TUG (TUG=Timed up and go)

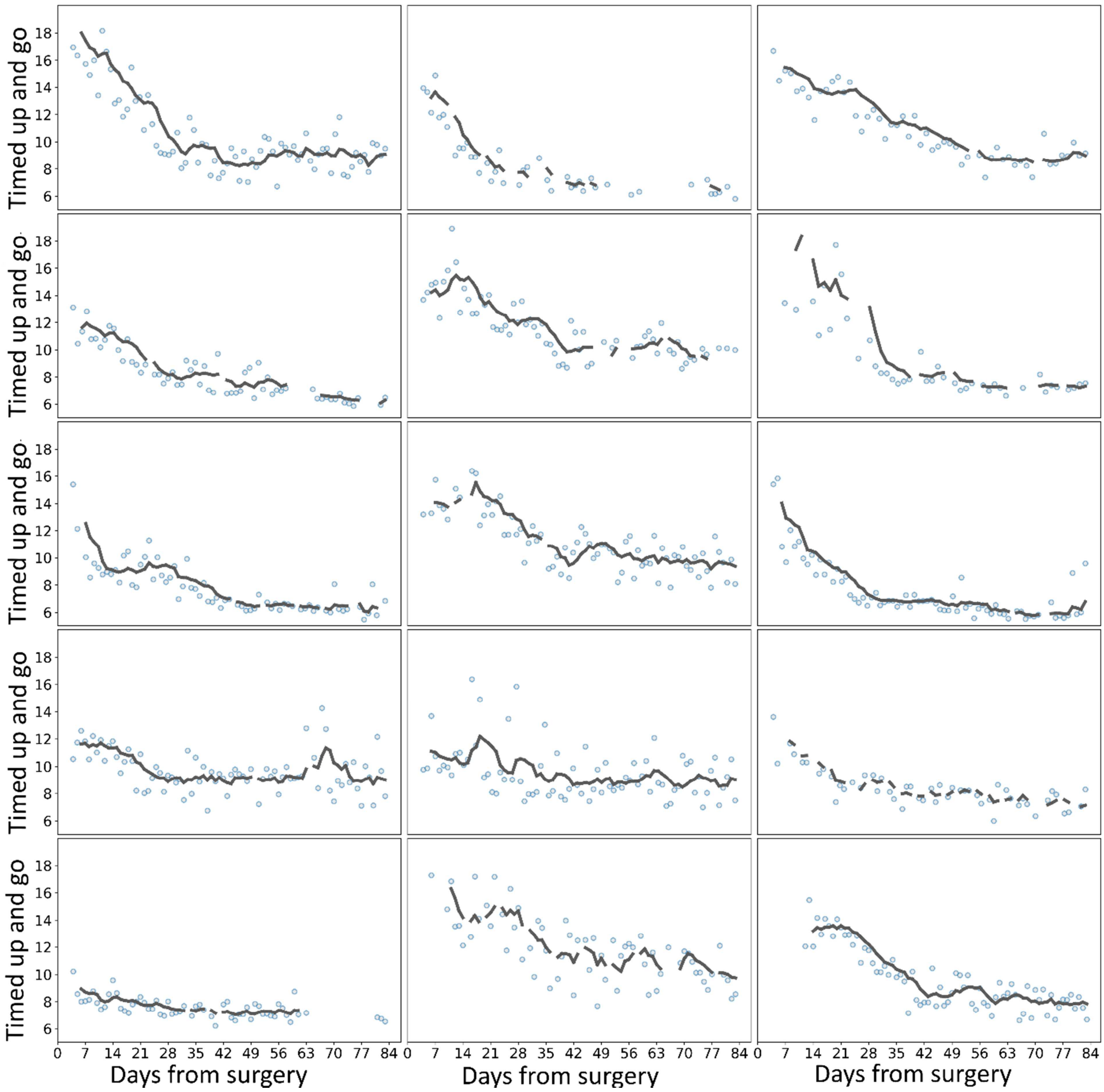

Figure 4. Remote timed up and go (TUG) estimated from activities of daily living measured by senior mobility monitor (black lines). Each panel represent remote TUG estimates one patient. Points represent remote TUG estimate obtained using one day of sensor data; dashed line: 7-days sliding-window moving average of remote TUG, with at least 3 days with sufficient wearing time in each sliding-window.

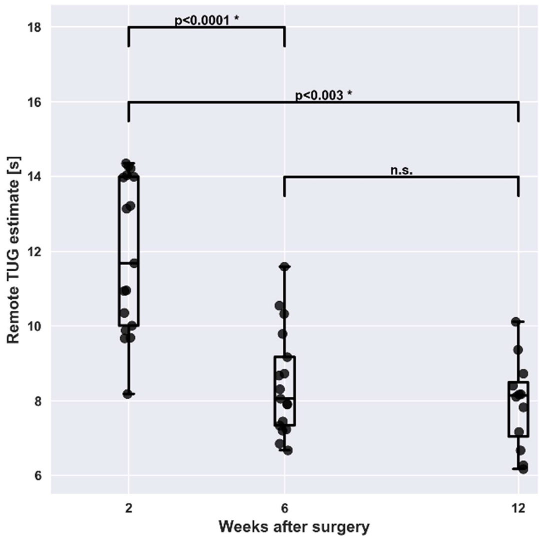

Figure 5. Remote timed up and go (TUG) estimated from activities of daily living measured by senior mobility monitor in a total-hip arthroplasty population. Predictions at weeks 2, 6, 12 after surgery are shown for the entire population. p-values refer to a paired t-test for the remote TUG at different weeks, after Bonferroni correction for multiple comparison (n=12 weeks, 121 total comparisons) n. s. = non-statistically significant.